A PROBABILISTIC APPROACH TO BLOCK SIZES IN RANDOM MAPS

LOUIGI ADDARIO-BERRY

ABSTRACT. We present a probabilistic approach to the core-size in random maps, which yields straightforward and singularity analysis-free proofs of some results of [2, 3, 7]. The proof also yields convergence in distribution of the rescaled size of the k'th largest 2-connected block in a large random map, for any fixed $k \geq 2$, to a Fréchet-type extreme order statistic. This seems to be a new result even when $k = 2$.

1. INTRODUCTION

The paper [2] is reasonably called the culmination of an extended line of research into core sizes in large random planar maps. The paper is an analytic *tour de force*, proceeding via singularity analysis of generating functions and the coalescing saddlepoint method. Banderier et al. [2] demonstrate how this powerful set of tools can be used to derive to local limit theorems and sharp upper and lower tail estimates. In particular, their theorems unify and strengthen the results from [3, 7].

The purpose of this note is to explain a probabilistic approach to the study of large blocks in large random maps. We end up proving two results. One is a weakening of [2, Theorem 7], the other a strengthening of [2, Proposition 5]. The main point, though, is that our approach, which is to reduce the problem to a question about outdegrees in conditioned Galton-Watson trees, feels direct and probabilistically natural (and short).

The remainder of the introduction lays out the definitions required for the remainder of the work. Section 2 explains how to view planar maps as composite structures. Randomness finally arrives in Section 3, which also contains the statements and proofs of this work's proposition, corollary, and theorem.
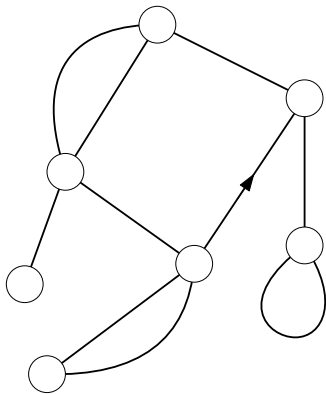
(A) A map M $= (M, \rho)$. The root edge ρ is drawn pointing from ρ^- to ρ^+.
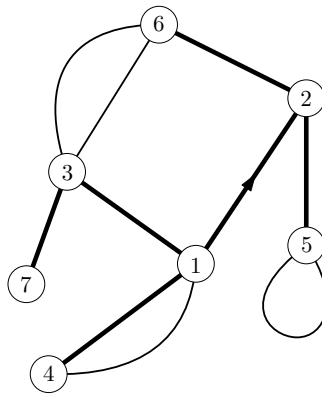
(B) The breadth-first search tree of M has bold edges. Vertices are labelled in increasing order according to $<_M$.
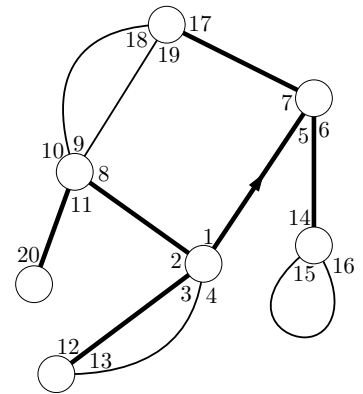
(C) Oriented edges/corners are labelled in increasing order according to $\prec_M$.

1.1. **Notation for maps and trees.** In this work, a rooted map is a pair $M = (M, \rho)$, where M is a connected combinatorial planar map[1] and $\rho = \rho^-\rho^+$ is an oriented edge of M with tail ρ^- and head ρ^+. We view M as embedded in $\mathbb{R}^2$ so that the unbounded face lies to the right of ρ; this in particular gives meaning to the "interior" and "exterior" for cycles of M (see Figure 1a). Write $v(M)$, $e(M)$ and $\bar{e}(M)$ for the vertices, edges and oriented edge of M, respectively. When convenient we write $v(M)$, etcetera, instead of $v(M)$. The *size* of a map is its number of edges; map M is larger than map M′ if $|e(M)| \geq |e(M')|$. The *trivial* map is the map with one vertex and no edges.

A *tree* is a connected rooted map $T = (T, \rho)$ with no cycles. We refer to ρ^- as the root of T. Children and parents are then defined in the usual way. The *outdegree* of $v \in v(T)$ is the number of children of v in T.

Given a map $M = (M, \rho)$, write $<_M$ for the total order of $v(M)$ induced by a breadth first search starting from ρ^- using the counterclockwise order of edges around a vertex to determine exploration priority (see Figure 1b). Listing the vertices according to this order as $v_1, v_2, \ldots, v_{|v(M)|}$, we in particular have $v_1 = \rho^-$, $v_2 = \rho^+$. We sometimes refer to $<_M$ as lexicographic order.
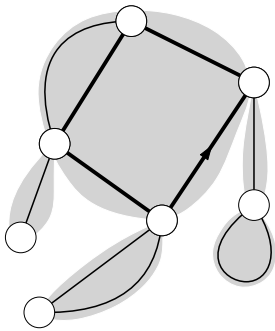

Breadth-first search builds a spanning tree $F = F(M)$ of M rooted at $v_1 = \rho^-$: for each $v \neq \rho^-$, the parent $p(v)$ of v in F is the $<_M$-minimal neighbour w of v_i. (There may be multiple edges of M joining a node w to a child v of w, but only one of these is an edge of F; here is how to determine which. If $w = \rho^- = v_1$ then take the first copy of each edge leaving w in counterclockwise order around w starting from $\rho = \rho^-\rho^+$. If $w \neq \rho^-$ then take the first copy of each edge leaving w in counterclockwise order starting from $wp(w)$; this makes sense inductively since $p(w) <_M w$.)

A *corner* of M is a pair (uv, uw) of oriented edges, where uw is the successor of uv in counterclockwise order around v. It is useful to identify oriented edges with corners: the corner corresponding to uv is the corner lying to the left of its tail. This is a bijective correspondence. We define a total order $\prec_M$ on the set of corners (equivalently, the set of oriented edges) of M as follows

FIGURE 1. Blocks are shaded; the root block has bold edges.

(see Figure 1c): say $uv \prec_M u'v'$ if either (a) $u <_M u'$ or (b) $u = u'$ and uv precedes $u'v'$ in counterclockwise order around u starting from $up(u)$ (or, if $u = v_1 = \rho^-$, starting from ρ).

2. PLANAR MAPS AS COMPOSITE STRUCTURES

We say a rooted map M is *separable* if there is a way to partition $e(M)$ into nonempty sets E and E' so that there is exactly one vertex v incident to edges of both E and E'. If M is not separable it is called 2-*connected*.[2] Write $\mathcal{M}$ for the set of rooted maps, and $\mathcal{B}$ for the set of 2-connected rooted maps.

This paragraph is illustrated in Figure 1. The maximal 2-connected submaps of M are called the *blocks* of M (hence the notation $\mathcal{B}$). They are edge-disjoint, and have a natural tree structure associated to them: the *block tree* of M has a node for each block of M, and two nodes are adjacent precisely if the corresponding blocks share a vertex. Write $B = B(M)$ for the maximal 2-connected submap of M containing ρ; call B the *root block*.

For each oriented edge uv of B, there is a (possibly empty) submap of M disjoint from B except at u and lying to the left of uv. We denote this map $M_{uv} = (M_{uv}, \rho_{uv})$,

---

[1]By "combinatorial" we mean that the map is uniquely specified by the cyclic ordering of edges around each vertex induced by the planar structure; equivalently, maps are defined "modulo orientation-preserving homeomorphisms of $\mathbb{R}^2$".

[2]The terminology of graphs and of maps are slightly at odds here. Many graph theorists would consider the "lollipop" graph, with one loop and one non-loop edge, to be 2-connected. As a map, it is not.

and call it the *pendant submap at uv* (or at the corresponding corner of B). Here ρ_{uv} is the edge of M following uv in counterclockwise order around u. We allow that M_{uv} is the trivial map. See Figure 2 for an illustration. We may reconstruct M from B and the $2|e(\mathrm{B})|$ submaps $\{\mathrm{M}_{uv}, \{u,v\} \in e(\mathrm{B})\}$ by identifying the tail of the root edge of M_{uv} with $u \in v(\mathrm{M})$ in such a way that the root edge of M_{uv} lies to the left of uv.

Compositionally, we thereby obtain that *rooted maps are 2-connected maps of rooted maps*. To formalize this, let $\mathcal{M}_n$ (resp. $\mathcal{C}_n$) be the set of rooted maps (resp. rooted 2-connected maps) with n edges, and write $M_n = |\mathcal{M}_n|$, $C_n = |\mathcal{C}_n|$. We take $C_0 = 1 = M_0$. Then with $M(z) = \sum_{n \geq 0} M_n z^n$ and $C(z) = \sum_{n \geq 0} C_n z^n$, we have

$$M(z) = C(zM(z)^2). \tag{1}$$

Now, introduce a formal variable y with $y^2 = z$. Then with $H(y) = yM(y^2) = z^{1/2}M(z)$, by (1) we have $h(y) = yC(h(y))$ so, by Lagrange inversion,

$$[z^n]M(z) = [y^{2n+1}]h(y) = \frac{1}{2n+1}[y^{2n}]C(y)^{2n+1}.$$

FIGURE 2. M_{uv} and M_{wu} are respectively dotted and dashed.

Here is the combinatorial interpretation of this identity. Given a map $\mathrm{M} = (M, \rho)$, represent the block structure of M by the following plane tree T_{M} defined as follows. (The construction is illustrated in Figure 3.) Let $\mathrm{B} = (B, \rho)$ be the block containing ρ, and list the *oriented* edges $\bar{e}(B)$ according to the order $\prec_{\mathrm{B}}$ as $a_1, \ldots, a_{2|e(B)|}$. We say that the root $\emptyset$ of T_{M} *represents* B in T_{M}.

The node $\emptyset$ has $2|e(B)|$ children in T_{M}. List them from left to right as $1, \ldots, 2|e(B)|$. Fix $i \in \{1, \ldots, 2|e(B)|\}$. If the counterclockwise successor $e_i = e_i^- e_i^+$ of a_i around a_i^- in M is also in $\bar{e}(B)$ then the corner formed by a_i and e_i contains no pendant submap. In this case i is a leaf in T_{M}. Otherwise, $e_i \in \bar{e}(M) \setminus \bar{e}(B)$. In this case write M_i for the connected component of $(v(M), e(M) \setminus e(B))$ containing $\{e_i^-, e_i^+\}$, and let $\mathrm{M}_i = (M_i, e_i)$. The subtree of T_{M} rooted at i is recursively defined to be the tree T_{M_i}. Figure 3a and 3c show a map M and a schematic representation of its blocktree. Figure 3b shows the corresponding tree T_{M}.

If M is 2-connected then T_{M} is simply a root of outdegree $2|e(M)|$ whose children are all leaves. More generally, for each block B of M, there is a corresponding node of T_{M} with exactly $2|e(B)|$ children. In other words, given the tree T_{M}, the block sizes in M are known.

Given the map B_ρ, the map M may be reconstructed by identifying e_i^- (the tail of the root edge of $\mathrm{M}_i = (M_i, e_i)$) and a_i^- so that e_i follows a_i in counterclockwise order around a_i^-. (This was explained in the paragraph preceding (1).) It follows recursively that M is uniquely specified by T_{M} together with the set of maps $(\mathrm{B}_v, v \in v(T_{\mathrm{M}}))$, where B_v is the block of M represented by v in T_{M}. If v is a leaf, take B_v to be the trivial map. Note that every node v has precisely $2|e(\mathrm{B}_v)|$ children in T_{M}. For the map M from Figure 3a, the nontrivial blocks represented by nodes of T_{M} are shown with identifying labels in Figure 3d.
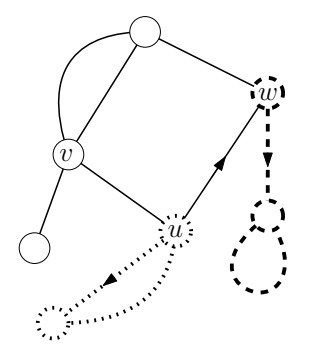
3. RANDOM MAPS

Let $\mathrm{M}_n \in_u \mathcal{M}_n$; this notation means that M_n is a random variable uniformly distributed over the (finite) set $\mathcal{M}_n$. We now describe the law of the tree T_{M_n}. Recall that $M_n = |\mathcal{M}_n|$ and $C_n = |\mathcal{C}_n|$, and that

$$M_n = \frac{2 \cdot 3^n (2n)!}{(n+2)!n!}.$$

(A) A map M.

(B) The tree T_M. Tiny squares represent trivial blocks.

(C) The decomposition of M into blocks. Blocks are joined by grey lines according to the block tree structure. Root edges of blocks are shown with arrows.

(D) The correspondence between blocks and nodes of T_M. Non-trivial blocks receive the alphabetical label (from A through L) of the corresponding node.
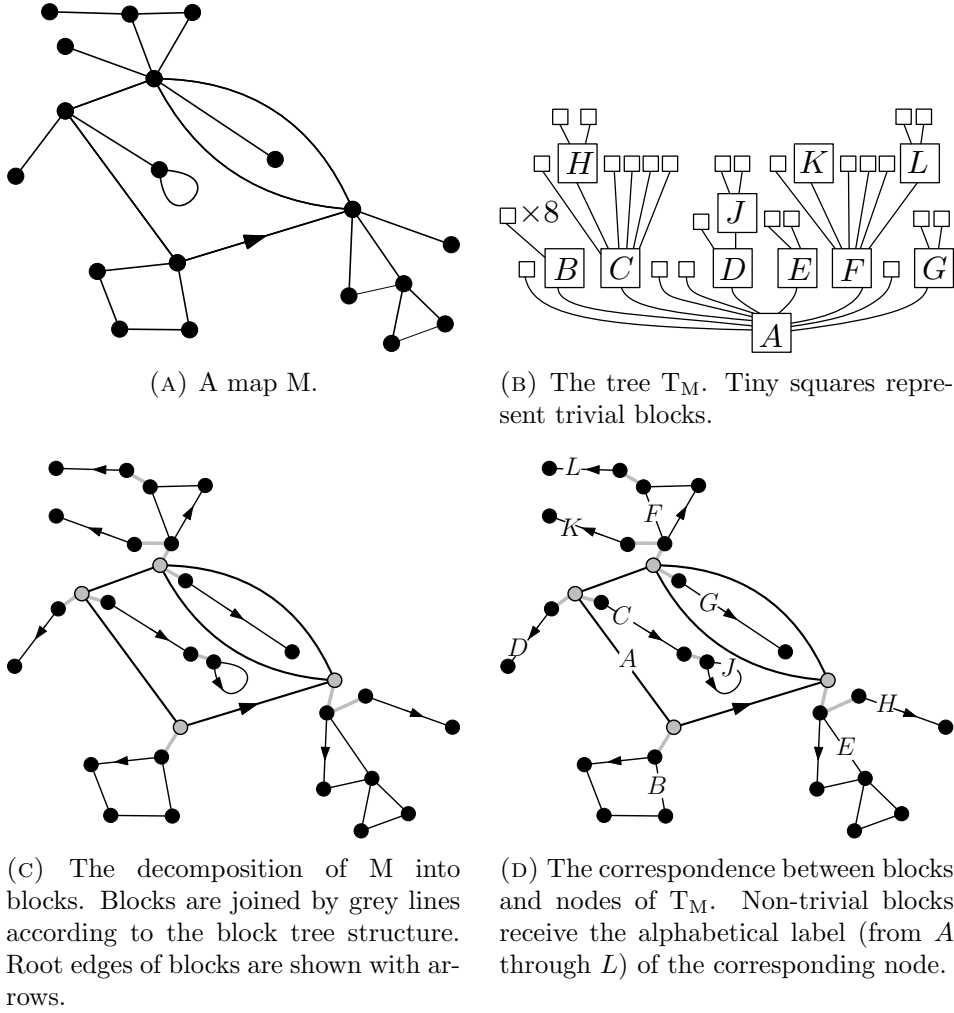
FIGURE 3. The relation between a map M and the plane tree T_M.

Using this, the compositional equation (1), and a little thought (see [8], pages 152-153), Lagrange inversion yields

$$C_0 = 1, \quad C_k = \frac{2(3k-3)!}{k!(2k-1)!} \quad \text{for } k \geq 1. \tag{2}$$

(The formulas for M_n and C_n are due to Tutte [12]; see also Brown [4].) Using Stirling's approximation, the formula (2) for $C_k = |\mathcal{C}_k|$ implies that $C(z)$ has radius of convergence $4/27$. Furthermore, it is straightforward to calculate that $C(4/27) = 4/3$, and that $\widehat{C}(4/27) := \sum_{k \geq 0} k(4/27)^k \cdot C_k = 4/9$. The fact $C(4/27)$ is finite is used straightaway; the second identity is noted for later use.

Fix $z \in (0, 27/4]$ and define a probability measure μ^z on $\mathbb{N}$ by

$$\mu^z(\{2k\}) = \frac{C_k z^k}{C(z)}.$$

Let T^z be a Galton-Watson tree with offspring distribution μ^z, and let T_n^z be a random tree whose law is that of T^z conditional on $|e(T^z)| = 2n$.

Proposition 1. *For all $z \in (0, 27/4]$, the trees T_n^z and T_{M_n} have the same law.*

Proof of Proposition 1. Fix a rooted tree t with $2n$ nodes, and list their outdegrees in lexicographic order as $d_1, \ldots, d_{2n}$; we assume all these are even. We saw in Section 2 that

a map M is uniquely specified by the tree T_{M} together with 2-connected maps $(\mathrm{B}_i, 1 \le i \le 2n)$, where B_i has d_i edges. It follows that the number of maps M with $T_{\mathrm{M}} = t$ is precisely

$$m(t) = \prod_{i=1}^{2n} \frac{C_{d_i}}{2}.$$

Therefore, $\mathbf{P}\{T_{\mathrm{M}_n} = t\}$ is proportional to $m(t)$. It is easily seen that this is also true for $\mathbf{P}\{T_n^z = t\}$ whatever the value of $z \in (0, 27/4]$. □

For the remainder of the section, let $(X_i, i \ge 1)$ be iid with law μ, and write $S_k = \sum_{i=1}^{k} X_i$. Now write $\mu = \mu^{4/27}$ and $T_n = T_n^{27/4}$.

Corollary 2. *List the outdegrees in T_n as in lexicographic order as $D_1, \ldots, D_{2n}$, and let σ be a uniformly random cyclic permutation of $\{1, \ldots, 2n\}$. Then the conditional law of $(X_1, \ldots, X_{2n})$ given that $S_{2n} = 2n - 1$ is precisely that of $(D_{\sigma(1)}, \ldots, D_{\sigma(2n)})$.*

Proof. This follows immediately from Proposition 1 and the cycle lemma Pitman [11, Lemma 6.1]. □

The corollary allows statistics about block sizes in M_n to deduced by studying a sequence of IID random variables conditioned on its sum. Pitman [11] explains a quite general link between probabilistic analysis of composite structures and randomly stopped sums; he calls this *Kolchin's representation of Gibbs partitions*. In a sense, the point of this note is to place the study of block sizes in maps within the latter framework.

We now state our main and only theorem. Let A be a Stable$(3/2)$ random variable, characterized by its Laplace transform:

$$\mathbf{E}\left[e^{-tA}\right] = e^{\Gamma(-3/2)t^{3/2}} = e^{(4\pi^{1/2}/3)t^{3/2}};$$

and for $k \ge 2$ let G_k be distributed as a standard $\Gamma(k-1)$ random variable.

Theorem 3. *Let $\mathrm{M}_n \in_u \mathcal{M}_n$, and for $k \ge 1$ let $L_{n,k}$ be the number of edges in the k'th largest block of M_n. Then as $n \to \infty$,*

$$\frac{L_{n,1} - n/3}{(8/(27\pi))^{1/2}n^{2/3}} \xrightarrow{\mathrm{d}} A,$$

and for any $k \ge 2$,

$$\frac{L_{n,k}}{(2\pi/3)^{1/3}n^{2/3}} \xrightarrow{\mathrm{d}} G_k^{-3/2}.$$

Remarks

(1) The second statment – the convergence of $L_{n,k}$ after rescaling when $k \ge 2$ – seems to be new. The fact that $(L_{n,2}/n^{2/3}, n \ge 1)$ is a tight family of random variables, or in other words that the second largest block has size $O(n^{2/3})$ in probability, is proved in [7] in some cases, and in [2] in greater generality.

(2) The convergence of $L_{n,1}$ is related to results from [3] and [7]. A stronger, local limit theorem for $L_{n,1}$, with explicit estimates on the rate of convergence, is given in Banderier et al. [2, Theorem 3]. As mentioned earlier, the initial motivation for the current work was to show how results in this direction may be straightforwardly obtained by probabilistic arguments. With a little care, the definition of the block tree may be altered to accommodate any of the compositional schemas considered in [2].

(3) In view of the preceding comment, the same line of argument should yield a version of the theorem (with constants altered appropriately) corresponding to any reasonable decomposition of a map into submaps of higher connectivity. Indeed, it seems that composite structures should in general fit within the current analytic

framework. (Of course, the sorts of limit theorems one may expect will depend on the combinatorics of the specific problem. As far as I am aware, the fact that the combinatorics of maps always lead to $O(n^{2/3})$ fluctuations and Airy-type limits is thus far an empirical fact rather than a provable necessity.)

(4) The paper [9] is a 150-page beast[3], so one may reasonably be skeptical that a proof which relies upon it can be called a "simplification" of anything. Here are two responses. First, the result from [9] that we use, namely Theorem 19.34, is self-contained; its proof is only 4 pages long, is elementary and probabilistic, and does not rely on results from elsewhere in the paper. Second, our reference to [9] could be removed by appealing to Theorem 1 of [1] (using (2.7) from[1] to control $L_{n,1}$); its (probabilistic) proof totals under two pages. (However, the language in [9] is closer to that of the current paper, which makes it slightly easier to apply.)

Proof of Theorem 3. List the blocks of M_n in decreasing order of size (number of edges) as $C_1, \ldots, C_K$, breaking ties arbitrarily. By Proposition 1, the sequence $(2e(C_i), 1 \leq i \leq k)$ has the same law as the decreasing rearrangement of non-zero outdegrees in T_n.

Let $X^{(1)}, \ldots, X^{(2n)}$ be the decreasing rearrangement of $X_1, \ldots, X_{2n}$. By Corollary 2, it follows that for all i we have

$$\mathbf{P}\left\{|e(C_1)| = i\right\} = \mathbf{P}\left\{X^{(1)} = 2i \mid S_{2n} = 2n - 1\right\}, \quad \text{and}$$
$$\mathbf{P}\left\{|e(C_2)| = i\right\} = \mathbf{P}\left\{X^{(2)} = 2i \mid S_{2n} = 2n - 1\right\}. \tag{3}$$

The largest values of such collections of random variables have been studied in detail by Janson [9]. Many of the results are phrased in terms of statistics of random balls-into-boxes configurations; the connection between this and outdegrees in conditioned Galton-Watson trees is made explicit in Section 8 of [9]. One of the themes running through that work is that of *condensation*: for heavy-tailed random variables, conditioning a sum S_m to be large is often equivalent to conditioning on having a single exceptionally large summand; furthermore, once the largest summand is removed, the remaining $m - 1$ summands are asymptotically distributed as iid random variables with their original distribution. See [1, 6, 10] for other instances of this phenomenon in related settings.

In the setting of this paper, Janson [9, Theorem 19.34 (iv), (vi)] provides local limit theorems for the sizes of $X^{(k)}$, for any $k \geq 1$. Verifying the conditions of that theorem are straightforward. First, the values of $C(4/27)$ and $\widehat{C}(4/27)$ imply that $\sum_{j \geq 0} 2j\mu(\{2j\}) = 2/3$. Furthermore, as $j \to \infty$, by Stirling's formula we have

$$\mu(\{2j\}) \sim \left(\frac{8}{27\pi}\right)^{1/2} j^{-5/2}.$$

Finally, the moment generating function of X clearly has radius of convergence 1. In the notation of [9], this says that $\nu = 2/3$, $\beta = 5/2 = \alpha + 1$, $\lambda = 1$, and $c = c' = (8/(27\pi))^{1/2}$. Using the identities from (3), Parts (iv) and (vi) of [9, Theorem 19.34], respectively, then give that

$$\frac{|e(C_1) - n/3|}{cn^{2/3}} \xrightarrow{\text{d}} A,$$

and for $k \geq 1$,

$$\frac{|e(C_k)|}{(3c/2)^{2/3}n^{2/3}} \xrightarrow{\text{d}} G_k^{-3/2}.$$

In view of the explicit expression for c, this proves the theorem. $\qquad\square$

---

[3]It is also a beauty.

Here are two final remarks. First, as mentioned above, the paper [2] proves a local limit theorem for $L_{n,1}$, with explicit error bounds in the rate of convergence. It would be interesting to recover such bounds by probabilistic methods. Second, that paper also proves essentially sharp bounds for the upper and lower tail probabilities of $L_{n,1}$; see Theorems 1 and 5. Similar tail bounds should apply in the more general settings of [1, 9]. This seems like a fundamental question in large deviations of functions of iid random variables. The main result of [5] seems quite pertinent, but pertains specifically to sums rather than to more general functions.

4. ACKNOWLEDGEMENTS

REFERENCES

[1] Inés Armendáriz and Michail Loulakis. Conditional distribution of heavy tailed random variables on large deviations of their sum. *Stochastic Process. Appl.*, 121(5):1138–1147, 2011. URL http://arxiv.org/abs/0912.1516.

[2] Cyril Banderier, Philippe Flajolet, Gilles Schaeffer, and Michèle Soria. Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Structures Algorithms*, 19(3-4): 194–246, 2001. URL http://algo.inria.fr/flajolet/Publications/BaFlScSo01.pdf.

[3] Edward A. Bender, L. Bruce Richmond, and Nicholas C. Wormald. Largest 4-connected components of 3-connected planar triangulations. *Random Structures Algorithms*, 7(4):273–285, 1995.

[4] William G. Brown. Enumeration of non-separable planar maps. *Canad. J. Math.*, 15:526–545, 1963. URL http://cms.math.ca/cjm/a145389.

[5] D. Denisov, A. B. Dieker, and V. Shneer. Large deviations for random walks under subexponentiality: the big-jump domain. *Ann. Probab.*, 36(5):1946–1991, 2008. URL http://dx.doi.org/10.1214/07-AOP382.

[6] Pablo A. Ferrari, Claudio Landim, and Valentin V. Sisko. Condensation for a fixed number of independent random variables. *J. Stat. Phys.*, 128(5):1153–1158, 2007. URL http://arxiv.org/abs/math/0612856.

[7] Zhicheng Gao and Nicholas C. Wormald. The size of the largest components in random planar maps. *SIAM J. Discrete Math.*, 12(2):217–228, 1999.

[8] Ian P. Goulden and David M. Jackson. *Combinatorial enumeration.* Dover Publications, Inc., Mineola, NY, 2004.

[9] Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012. URL http://dx.doi.org/10.1214/11-PS188.

[10] I. Kortchemski. Limit theorems for conditioned non-generic galton-watson trees. *Ann. Inst. H. Poincaré Probab. Statist.*, 2015 (to appear). http://arxiv.org/abs/1205.3145.

[11] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. URL http://www.stat.berkeley.edu/~pitman/621.pdf.

[12] W. T. Tutte. A census of planar maps. *Canad. J. Math.*, 15:249–271, 1963. URL https://cms.math.ca/10.4153/CJM-1963-029-x.

DEPARTMENT OF MATHEMATICS AND STATISTICS, MCGILL UNIVERSITY, 805 SHERBROOKE STREET WEST, MONTRÉAL, QUÉBEC, H3A 2K6, CANADA

E-mail address: louigi.addario@mcgill.ca

URL: http://www.problab.ca/louigi/